

(Por) **Identificação de Clusters de População Não Servidos por Transporte Público**

Objetivo:

O objetivo deste teste é criar uma representação visual ou geométrica da área de influência de cada parada de transporte público, considerando apenas sua existência física. Na realidade, este produto teria como finalidade facilitar a análise posterior da cobertura do transporte público em relação à população, representada por pontos, em uma determinada região. Além disso, busca-se identificar clusters de população que não estão sendo atendidos pelo sistema de transporte público, permitindo uma compreensão mais detalhada das áreas onde há uma lacuna na cobertura do transporte público.

Dados Disponíveis:

- Pontos de População: Uma base de dados contendo coordenadas geográficas (latitude e longitude) dos pontos de população na região de interesse.
- GTFS (General Transit Feed Specification): Um conjunto de dados estruturado contendo informações sobre as paradas de transporte público, incluindo coordenadas geográficas (latitude e longitude) e outros atributos relevantes.

Tarefas:

1) Carregar os dados:

- Importar as bibliotecas necessárias em Python.
- Carregar os dados dos pontos de população e as informações GTFS das paradas de transporte público.

2) Preparar os dados:

- Verificar a consistência e integridade dos dados.
- Realizar a limpeza dos dados, se necessário, remover ou corrigir entradas dupas e/ou inconsistentes.
- Garantir que os dados estejam em um formato adequado para análise posterior.

3) Visualizar os dados espacialmente:

- Plotar os pontos de população em um mapa para visualizar sua distribuição geográfica.
- Plotar as paradas de transporte público em um mapa para identificar sua localização.
- Visualizar a sobreposição dos pontos de população e as paradas de transporte público.

4) Criar área/geometria de alcance das paradas:

- Criar áreas de alcance que representem a proximidade de cada parada (por exemplo, um raio de 500m, usar alguma ferramenta opensource para identificar a área de alcance a pé das paradas).
- Visualizar a sobreposição ou a sua falta entre o alcance das paradas e os pontos de população.

5) Verificar quais os pontos de população estão fora das áreas de alcance de cada parada:

- Utilizar as áreas de alcance das paradas para identificar os pontos de população que estão descobertos.
- Realizar análise descritiva simples do resultado de população coberta e descoberta.

6) Identificar os clusters de população não servidos pelo transporte público:

- Utilizar técnicas de análise espacial ou agrupamento para identificar agrupamentos (clusters) de pontos de população que não estão sendo atendidos pelo sistema de transporte público.
- Realizar análise descritiva simples dos clusters identificados.

Entregáveis:

- Espera-se que o candidato forneça um script em Python.
- Um arquivo contendo a(s) solução(s) visual(s), que pode ser em formato raster ou vetor.
- Fornecer um documento de explicação textual (.txt ou .pdf).

(ENG) **Population Clusters Identification Not Served by Public Transportation**

Objective:

The objective of this test is to create a visual or geometric representation of the influence area of each public transport stop, considering only its physical existence. In reality, this product would aim to facilitate the subsequent analysis of public transport coverage in relation to the population, represented by points, in a given region. Additionally, it seeks to identify clusters of population that are not being served by the public transport system, allowing for a more detailed understanding of areas where there is a gap in public transport coverage.

Available Data:

- Population Points: A database containing geographic coordinates (latitude and longitude) of population points in the area of interest.
- GTFS (General Transit Feed Specification) - Public Transportation Stops: A structured dataset containing information about public transportation stops, including geographic coordinates (latitude and longitude) and other relevant attributes.

Tasks:

1) Load the data:

- Import the necessary libraries in Python.
- Load the population points data and GTFS information of public transportation stops.

2) Prepare the data:

- Check data consistency and integrity.
- Perform data cleaning, if necessary, by removing or correcting duplicate and/or inconsistent entries.
- Ensure data is in a suitable format for further analysis.

3) Visualize the data spatially:

- Plot population points on a map to visualize their geographic distribution.
- Plot public transportation stops on a map to identify their location.
- Visualize the overlap between population points and public transportation stops.

4) Create stop reach area/geometry:

- Create reach areas representing the proximity of each stop (e.g., a 500-meter radius, using an open-source tool to identify stop reach area based on walking distance).
- Visualize the overlap or lack of overlap between stop reach areas and population points.

5) Identify population points outside the reach areas of each stop:

- Use stop reach areas to identify population points outside these areas.
- Conduct simple descriptive analysis of the result of covered and uncovered population.

6) Identify clusters of population not served by public transportation:

- Use spatial analysis or clustering techniques to identify clusters of population points not being served by the public transportation system.
- Conduct simple descriptive analysis of the identified clusters.

Deliverables:

- The candidate is expected to provide a Python script.
- A file containing the visual solution(s), which could be in raster or vector format.
- Provide a textual explanation document (.txt or .pdf).